MOLECULAR AND CELLULAR BIOLOGY, Sept. 1986, p. 3156–3165
0270-7306/86/093156-10$02.00/0

Vol. 6, No. 9

Structure, Organization, and Sequence of Alpha Satellite DNA from Human Chromosome 17: Evidence for Evolution by Unequal Crossing-Over and an Ancestral Pentamer Repeat Shared with the Human X Chromosome

JOHN S. WAYE AND HUNTINGTON F. WILLARD*

Department of Medical Genetics, University of Toronto, Toronto, Ontario, Canada M5S 1A8

Received 1 April 1986/Accepted 19 May 1986

The centromeric regions of all human chromosomes are characterized by distinct subsets of a diverse tandemly repeated DNA family, alpha satellite. On human chromosome 17, the predominant form of alpha satellite is a 2.7-kilobase-pair higher-order repeat unit consisting of 16 alphoid monomers. We present the complete nucleotide sequence of the 16-monomer repeat, which is present in 500 to 1,000 copies per chromosome 17, as well as that of a less abundant 15-monomer repeat, also from chromosome 17. These repeat units were ~98% identical in sequence, differing by the exclusion of precisely 1 monomer from the 15-monomer repeat. Homologous unequal crossing-over is suggested as a probable mechanism by which the different repeat lengths on chromosome 17 were generated, and the putative site of such a recombination event is identified. The monomer organization of the chromosome 17 higher-order repeat unit is based, in part, on tandemly repeated pentamers. A similar pentameric suborganization has been previously demonstrated for alpha satellite of the human X chromosome. Despite the organizational similarities, substantial sequence divergence distinguishes these subsets. Hybridization experiments indicate that the chromosome 17 and X subsets are more similar to each other than to the subsets found on several other human chromosomes. We suggest that the chromosome 17 and X alpha satellite subsets may be related components of a larger alphoid subfamily which have evolved from a common ancestral repeat into the contemporary chromosome-specific subsets.

Alpha satellite (alphoid DNA) is a complex family of tandemly repeated DNA found in primate genomes. Long tandem arrays of alpha satellite DNA based on a monomer repeat length of ~171 base pairs (bp) are located principally at the centromeres of primate chromosomes (3, 15, 16, 19, 20). In the human genome, these sequences have been identified at the centromeric regions of each human chromosome and constitute some 5% of total human DNA. Human alpha satellite was initially described as a 340-bp *Eco*RI repeat, comprising two diverged monomer halves of 169 and 171 bp (16, 34). However, more recent studies have indicated that this DNA family is substantially more heterogeneous than was once believed. The dimeric *Eco*RI repeat itself is made up of several distinct subfamilies (11, 22), and a number of different molecular configurations in the human genome have been described (7, 10, 18, 28, 29, 32, 33, 35). Alpha satellite subsets have been identified on several human chromosomes, and the suggestion has been made that individual human chromosomes may each be characterized by distinct alpha satellite subsets defined by restriction enzyme periodicity and primary sequence (15, 18, 29).

Alpha satellite DNA specific for the human X chromosome (alphaX) has been cloned and extensively characterized (28, 32, 35). The 2.0-kilobase-pair (kb) *Bam*HI higher-order repeat unit from this chromosome consists of 12 tandem but diverged alpha satellite monomers arranged as two adjacent and related pentamer blocks plus an additional two monomers also related to monomers within the pentamer blocks (28). Under conditions which allow hybridization between heterologous chromosomal subsets, the 2.0-kb alphaX repeat detects alpha satellite domains on all human chromosomes (29). This cross-reaction between diverged subsets of this DNA family has been exploited to isolate representatives of an alpha satellite subset from the human Y chromosome (alphaY) (33). The Y-linked higher-order repeat unit is some 5.5 kb in length and is significantly diverged from both the X-linked 2.0-kb alpha satellite repeat and the simpler 340-bp alpha dimers (33).

In this report, we have used the alphaX repeat to identify and clone representatives of the alpha satellite family from human chromosome 17 (alpha17). The most prominent chromosome 17 higher-order repeat length is 2.7 kb, consisting of 16 tandem monomers. We have cloned copies of the 16-monomer higher-order repeat unit (16-mer), as well as copies of a less abundant 15-monomer (15-mer) repeat. These clones hybridize predominantly, if not exclusively, to chromosome 17 under conditions of high stringency. We have determined the complete nucleotide sequence of the alpha17 16-mer and 15-mer higher-order repeat units. These are shown to be essentially identical, with the exclusion of precisely 1 monomer from the 15-mer relative to the 16-mer. Homologous unequal crossing-over (13, 23) is suggested as a mechanism for the genesis of the different repeat lengths observed on chromosome 17, and the presumed site of such a recombination event is identified. Finally, the organization of monomers within the alpha17 repeats is shown to have common features with that of alphaX, indicating that these chromosomal subsets may have evolved from a common ancestral alphoid form.

MATERIALS AND METHODS

Human and somatic cell hybrid DNA. DNA was isolated from peripheral blood leukocytes, from cultured lymphoblast cells, and from somatic cell hybrid lines as described

* Corresponding author.

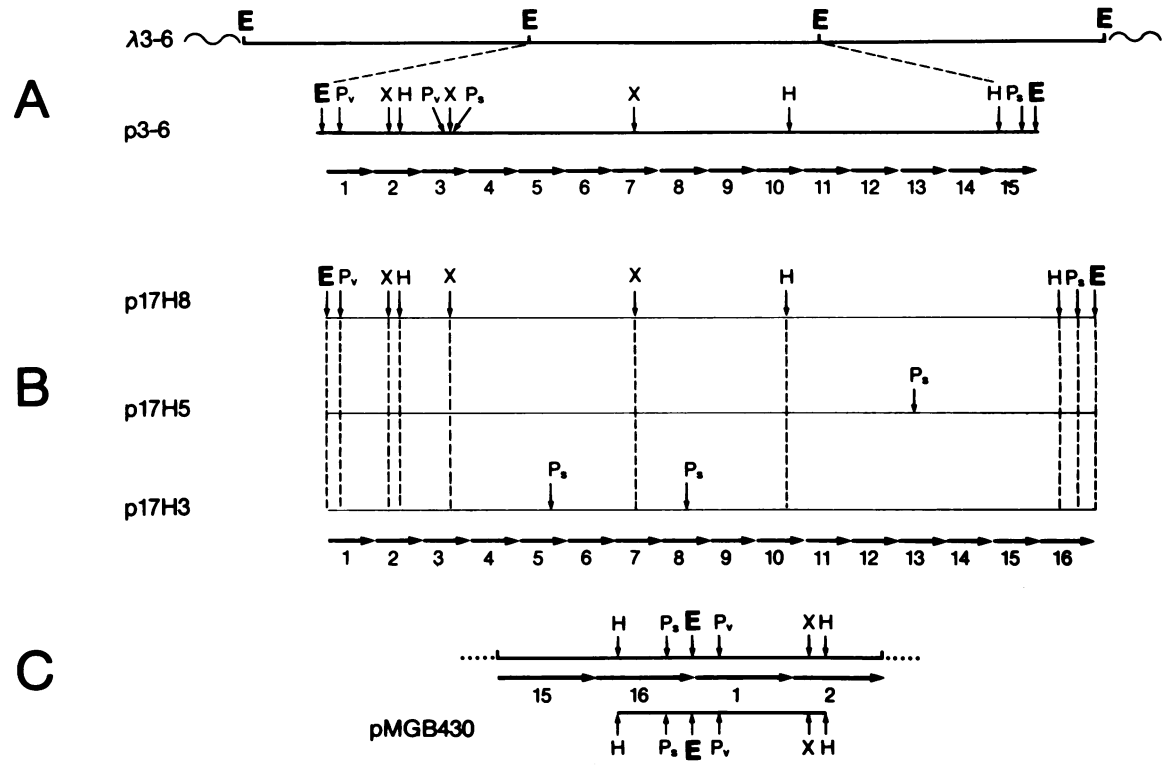

FIG. 1. Restriction maps and monomer organization of the alpha17 clones λ3-6 and p3-6 (A); p17H8, p17H5, and p17H3 (B); and pMGB430 (C). The *Eco*RI sites (E) which define the higher-order repeat units are shown in boldface type. Additional restriction sites shown are *Pst*I (Ps), *Pvu*II (Pv), *Xba*I (X), and *Hin*dIII (H). Vertical broken lines extending from the restriction sites of p17H8 (B) indicate that these sites are also present in p17H3 and p17H5. Tandemly arranged arrows represent monomer units (~171 bp) as determined by nucleotide sequencing. In panel C, the predicted structure of a tandem 16-mer junction fragment is shown above the restriction map of pMGB430.

(29). The somatic cell hybrids which contain single human chromosomes have been described (29), as have those containing multiple human chromosomes which form the mapping panel used in this study (30, 31).

Southern hybridization. Methods for restriction endonuclease digestion, electrophoresis, transfer to nitrocellulose, prehybridization, and hybridization have been described (28, 29, 32). Filters were washed as described previously (28, 32), either at reduced stringency (final wash in 0.5 M NaCl at 65°C) or at high stringency (final wash in 0.1× SSC [1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate] at 68 to 70°C).

Nucleotide sequencing. To facilitate rapid nucleotide sequencing, sets of deletion plasmids were constructed from p17H8 (2.7-kb *Eco*RI insert in vector pSP64) and p3-6 (2.5-kb *Eco*RI insert in pUC9) by using a strategy similar to that of Henikoff (9). The plasmids were linearized with *Bam*HI, which cuts only in the polylinkers of pSP64 (17) and pUC9 (27) and does not cut the alpha satellite inserts. Exonuclease digestions were done under conditions which gave a unidirectional digestion rate of ~100 bp/min (2 μg of linearized plasmid digested at 37°C with 250 U of exonuclease III in a buffer containing 6.6 mM Tris, pH 7.4, 6.6 mM $MgCl_2$, 6.6 mM 2-mercaptoethanol, and 60 mM NaCl). Portions were removed at 2-min intervals, and the reactions were stopped by adding an equal volume of 2× nuclease S1 buffer (60 mM sodium acetate, pH 4.5, 1 mM $ZnSO_4$, 0.2 mM NaCl, and 0.2% sodium dodecyl sulfate [SDS]). Single-stranded tails were removed by digestion with nuclease S1 (50 U per reaction at 37°C for 15 min). Samples were phenol extracted, ethanol precipitated, and digested with *Eco*RI. The p17H8 and p3-6 samples were then ligated into *Eco*RI-*Sma*I digested pSP64 and pUC9, respectively. This step favored the cloning of deleted satellite inserts because insertion of the deleted vectors would presumably form palindromes centered around the *Eco*RI site, structures which are thought to be incompatible with plasmid viability (8).

The p17H8 and p3-6 exonuclease III derivatives were sequenced by the dideoxytermination method (21) with ^{35}S-labeled triphosphate (1) and double-stranded plasmid templates as described by Korneluk et al. (12). The M13 reverse primer (New England Biolabs) was used to sequence the p3-6 derivatives. Sequencing of the p17H8 derivatives employed, in addition to the reverse primer, a primer complementary to a portion of the SP6 polymerase promoter of pSP64 (12). Deletion derivatives spanning both p17H8 and p3-6 (at 150- to 200-bp intervals) were sequenced, as were several restriction fragments conveniently subcloned from each. The entire p3-6 and p17H8 sequences were determined when possible on both strands. Otherwise, the sequence was determined on one strand with multiple overlapping deletion plasmids. pMGB430, a putative junction clone, was subcloned as two *Eco*RI-*Hin*dIII fragments into pSP64 and sequenced from both directions with SP6 and reverse primers.

RESULTS

Isolation of chromosome 17 alpha satellite clones. Using conditions that permit cross-hybridization with heterologous alpha satellite domains, we used the 2.0-kb alphaX probe pBamX7 (28, 32) to identify and isolate alpha satellite-containing clones from a library of human DNA constructed

with partial *Eco*RI digests in the lambda vector λgtWES.B (generously provided by K. E. Davies). One clone, λ3-6, contained three tandemly arranged 2.5-kb *Eco*RI fragments (Fig. 1A). One of these fragments was subcloned into pUC9, producing p3-6 which was examined in detail. When used as a probe in filter hybridization experiments, p3-6 hybridized with prominent bands of approximate lengths 2.7, 2.5, and 2.4 kb in *Eco*RI-digested human genomic DNA (Fig. 2A). Since the isolated clone contained representatives of only the 2.5-kb band and since preliminary mapping studies with somatic cell hybrid DNAs suggested that these *Eco*RI fragment lengths could be assigned to human chromosome 17 (see below), we used the p3-6 alpha satellite probe to identify and clone representatives of the more abundant 2.7-kb *Eco*RI repeat from chromosome 17. DNA from a mouse-human somatic cell hybrid containing only human chromosomes 5, 17, and 21 was digested to completion with *Eco*RI. The DNA was size-fractionated, and fragments 2.5 to 3.0 kb long were isolated and cloned into *Eco*RI-cleaved and alkaline phosphatase-treated pSP64. The resulting plasmid library, containing approximately 2,000 clones, was screened with the ^{32}P-labeled insert from p3-6 under highly stringent conditions to prevent potential cross-hybridization with alpha satellite from other human chromosomes (e.g., chromosome 5 or 21). Three clones (p17H3, p17H5, and p17H8) were isolated; their restriction maps are shown in Fig. 1B. The organization of alpha satellite monomers in the p3-6 and p17H clones is indicated in the figure, as determined by nucleotide sequence analysis (see below).

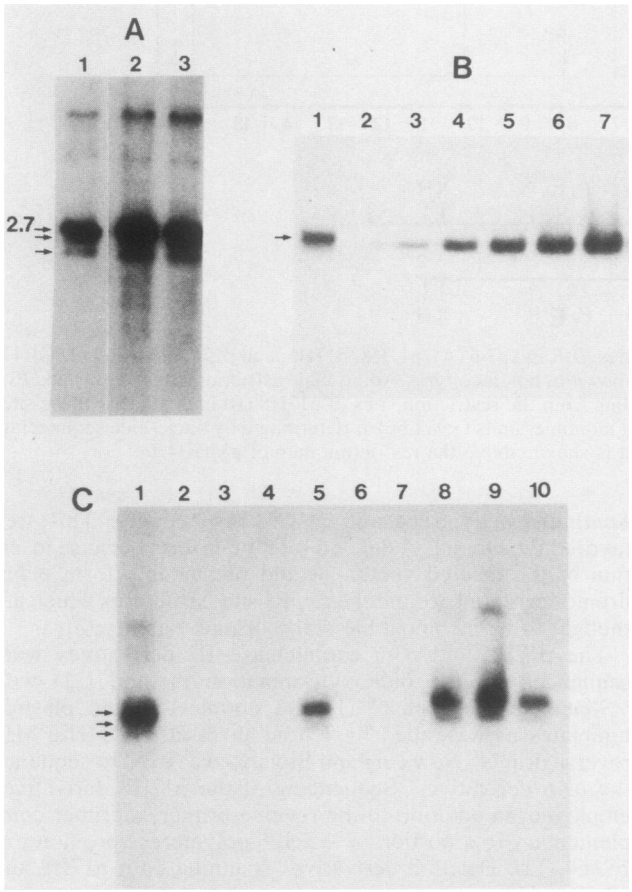

FIG. 2. Alpha17 genomic hybridization pattern, copy number, and chromosome mapping. (A) *Eco*RI digests of genomic DNA from three individuals. The three major hybridizing bands corresponding to 16-mer, 15-mer, and 14-mer lengths are indicated by arrows. The size of the 16-mer band is indicated in kilobases. (B) *Eco*RI-digested human genomic DNA and graded amounts of λ3-6. Lane 1, 5 μg of human genomic DNA; lanes 2 through 7, 5 μg of mouse DNA containing the equivalent of 0, 125, 250, 500, 1,000 and 2,000 copies, respectively, of the alpha17 repeat per haploid genome, calculated essentially as described (32). Identical results were obtained when p17H8 was used instead of λ3-6. The arrow indicates the 2.7-kb 16-mer fragment. (C) *Eco*RI-digested human DNA (lane 1), mouse DNA (lane 2), and DNA from somatic cell hybrids which either contain human chromosome 17 (lanes 5, 8, 9, 10) or have lost chromosome 17 (lanes 3, 4, 6, and 7). Arrows indicate the 16-mer, 15-mer, and 14-mer *Eco*RI fragments. For all panels, filter hybridization experiments were performed with p3-6 or p17H8 as probe. Filters were washed at high stringency.

TABLE 1. Chromosomal mapping of chromosome 17 alpha satellite DNA

Chromosome	Concordant[a]		Discordant[b]		% Discordant
	+/+	−/−	−/+	+/−	
1	4	13	9	3	41
2	5	15	8	1	31
3	6	12	7	4	38
4	5	11	8	5	45
5	9	10	4	6	34
6	5	16	8	0	28
7	7	9	6	7	45
8	2	12	11	4	52
9	0	15	13	1	48
10	3	9	10	7	59
11	4	10	9	6	52
12	5	11	8	5	45
13	7	14	6	2	28
14	8	7	5	9	48
15	3	14	10	2	41
16	4	13	9	3	41
17	13	16	0	0	0
18	4	9	9	7	55
19	3	10	10	6	55
20	9	9	4	7	38
21	7	8	6	8	48
22	4	13	9	3	41
X	11	2	2	14	55
Y	0	15	13	1	48

[a] Number of hybrids in which the particular human chromosome and the hybridizing bands were either both present (+/+) or both absent (−/−).
[b] Number of hybrids in which the human chromosome was absent and the hybridizing bands were present (−/+) or in which the human chromosome was present and the bands were absent (+/−).

Comparing the restriction maps of the 15-mer (p3-6) and 16-mer (p17H clones) higher-order repeat units, it is apparent that a single monomer length is missing from the 15-mer repeat relative to the 16-mer repeat and that the missing monomer lies between the *Hin*dIII sites in monomers 10 and 16 of p17H8 (Fig. 1B). The tandem arrangement of 15-mers in the genome is demonstrated directly by the λ3-6 isolate, containing three tandem 15-mers (Fig. 1A). (The observation that all three copies share the same 1-monomer deletion makes it highly unlikely that this deletion reflects a cloning artifact.) The tandem nature of 16-mers was suggested by partial digestion experiments, in which ladders of *Eco*RI- or *Pvu*II-digested fragments at 2.7-kb intervals could be detected in human genomic DNA (data not shown). To confirm this organization, a putative junction fragment between two tandem 16-mer *Eco*RI repeats was isolated from a library of *Hin*dIII-digested human DNA cloned in pUC9, with the p17H8 insert as probe. As shown in Fig. 1C, the clone pMGB430 has the predicted structure for such a junction

FIG. 3. Complete nucleotide sequence of p17H8. The 2,712-bp sequence is arranged as 16 contiguous monomers, with position 1 being the first base following the *Eco*RI cleavage site. The positions of base deletions in monomers are based entirely on the maximum homology alignment and are otherwise arbitrary. Base positions of each monomer are given at the right.

TABLE 2. Sequence comparisons among the 16 alpha17 monomers of p17H8[a]

Homology group and monomer no.	% Identity[b]															
	A 1	B 2	C 3	D 4	E 5	A 6	B 7	C 8	D 9	C 10	D 11	E 12	A 13	A 14	A 15	E 16
A 1	—	7	18	4	10	63	4	17	7	20	0	3	34	35	34	3
B 2	73	—	8	12	11	0	50	14	19	17	15	16	4	0	7	6
C 3	77	74	—	17	8	18	5	27	12	34	9	11	8	8	4	4
D 4	72	76	80	—	24	5	14	24	33	10	47	25	4	0	0	21
E 5	83	74	75	80	—	8	21	15	23	4	27	66	11	12	7	46
A 6	90	74	82	78	76	—	10	12	0	10	0	7	54	51	49	0
B 7	73	87	77	81	80	79	—	8	17	0	14	19	9	9	4	4
C 8	74	74	81	78	75	77	76	—	11	17	12	10	11	8	7	13
D 9	71	77	76	83	78	74	80	73	—	0	36	23	0	0	0	10
C 10	80	80	88	81	78	82	80	81	77	—	10	4	18	19	13	8
D 11	72	78	78	88	80	77	81	75	84	81	—	28	4	0	0	14
E 12	66	72	72	78	88	72	76	70	75	73	77	—	3	10	3	44
A 13	81	74	78	75	75	89	77	75	73	82	75	69	—	75	72	7
A 14	81	73	78	74	76	88	77	74	73	82	73	73	94	—	65	0
A 15	80	74	77	74	73	87	75	74	71	81	72	68	91	91	—	3
E 16	66	69	71	75	83	70	74	71	73	74	75	81	69	69	67	—

[a] Numbers on the axes refer to individual monomers (see Fig. 3). Letters on the axes refer to homology groupings defined in the text.
[b] Figures below the diagonal within the matrix represent the percent identity of the two compared monomers (100% = 171 bp). Figures above the diagonal represent comparison of nonconsensus positions only. See text for explanation.

fragment, as determined by restriction mapping and DNA sequence determination (see below).

Determination of chromosome location and number of copies of the higher-order repeat unit. To determine the location of the alpha satellite sequences homologous to the p3-6 or p17H8 clone, DNA from human cells, rodent cells, and a series of 29 rodent-human somatic cell hybrids (30, 31) was digested with *Eco*RI and examined in a filter hybridization experiment with either p3-6 or p17H8 as probe under high-stringency conditions (Fig. 2C). By correlating the presence or absence of the 2.7- to 2.4-kb *Eco*RI bands with the human chromosomal content of the hybrids, these bands could be assigned to human chromosome 17. All other human chromosomes were discordant in a high percentage of hybrids tested (Table 1). Thus, these clones belong to and define a subset of alpha satellite DNA from human chromosome 17 (alpha17).

To estimate the number of copies of the 16-mer and 15-mer higher-order alphoid repeats on chromosome 17, the hybridization signal obtained with known amounts of *Eco*RI-digested human DNA was compared with serial dilutions of similarly digested λ3-6 DNA (Fig. 2B) or p17H8 DNA (not shown). Based on these comparisons, we estimate that there are approximately 500 to 1,000 copies of the 16-mer repeat and approximately 100 copies of the 15-mer repeat per chromosome 17.

Nucleotide sequence of alpha17 clones. The complete nucleotide sequences of p3-6, p17H8, and pMGB430 have been determined. The sequence of p17H8 is shown arranged as 16 tandem monomers in Fig. 3. The sequences of p3-6 and pMGB430 are not shown here, but are available on request. Partial sequences of p17H3 and p17H5 (approximately 500 bp each) were also determined and were >99% identical to the sequence of p17H8. A similar degree of homology was noted for the putative junction clone pMGB430, supporting the conclusion that this clone represents a junction between 16-mers.

In contrast, p3-6 showed 2.2% sequence divergence compared with p17H8. The sequence differences were evenly distributed throughout the repeat. In addition, as predicted by consideration of the restriction map (Fig. 1A and B), p3-6 was missing exactly 1 monomer length (171 bp) relative to p17H8. The precise boundaries of the deletion cannot be determined because the missing monomer is from a region of the 16-mer repeat which has an apparent monomer triplication (see below). This results in there being three highly homologous (91 to 94%) monomers (numbers 13 through 15) in the 16-mer repeat and only two such monomers in the 15-mer repeat. The fact that a single monomer is absent from such a region is suggestive of homologous unequal crossing-over in the genesis of these repeat lengths. Such a mechanism could account for generation of the 15-mer from the 16-mer (or vice versa) and will be discussed in greater detail below.

Monomeric configuration. We have previously demonstrated that it is possible to derive the underlying monomeric configuration for a given higher-order tandem repeat unit by comparing homologies among its constituent monomers (28). Homology comparisons among monomers of p17H8 are given in Table 2. As performed previously for the human X chromosome (28), sequence similarities are expressed in two ways. First, direct sequence homologies are presented below the diagonal in Table 2. The 16 monomers are each 66 to 94% identical to each other. Second, a consensus identity index (28) is presented above the diagonal in Table 2. For this analysis, an alpha17 consensus sequence was determined for p17H8 by selecting the base present in a majority of the 16 monomers at each position. For comparisons between monomers, nucleotide positions in agreement with the consensus sequence were not considered (since they do not help in distinguishing one monomer from another); only variable, nonconsensus positions were compared between monomers. These nonconsensus positions were from 0 to 75% identical in pairwise comparisons among the 16 monomers (Table 2). For both sets of comparisons, it is apparent that the alpha17 monomers can be loosely arranged into five homology groupings, where monomers within a group are similar but not identical: group A (monomers 1, 6, 13, 14, and 15); group B (monomers 2 and 7); group C (monomers 3, 8, and 10); group D (monomers 4, 9, and 11); and group E (monomers 5, 12, and 16). (This grouping is most readily appreciated for the group A monomers 1, 6, 13, 14, and 15, each of which has a conserved 3-base gap at position 33–35 [Fig. 3].) Thus, the monomeric configuration of the 16

TABLE 3. Sequence comparisons between monomers of alpha17 and alphaX higher-order repeat units[a]

X chromosome monomers	Chromosome 17 monomers															
	A 1	B 2	C 3	D 4	E 5	A 6	B 7	C 8	D 9	C 10	D 11	E 12	A 13	A 14	A 15	E 16
C 1	73	73	80	80	72	80	75	75	73	84	77	71	75	75	74	68
D 2	75	78	82	91	81	80	82	81	84	82	89	77	76	75	74	77
E 3	64	66	67	72	81	68	69	65	72	67	72	82	64	65	64	77
A 4	79	72	76	74	73	85	75	72	73	76	73	69	81	81	80	69
B 5	69	86	75	78	77	74	86	74	78	78	78	75	73	72	73	74
C 6	77	75	93	81	76	81	77	83	76	89	79	70	78	78	77	71
D 7	68	73	75	85	78	73	76	73	79	77	84	74	70	70	68	73
E 8	67	68	71	75	82	72	73	71	74	74	75	82	70	71	68	80
A 9	78	72	74	72	73	83	75	70	71	76	72	70	81	79	79	70
B 10	72	82	75	78	77	78	87	75	80	78	81	73	77	76	74	71
C 11	73	70	78	77	74	76	73	87	71	81	75	67	75	74	73	70
D 12	71	80	75	84	80	75	81	73	89	78	84	75	74	74	73	75

[a] Figures represent percent identity of the two compared monomoers from the X chromosome and chromosome 17 (100% = 171 bp).

monomers in p17H8 can be represented as *Eco*RI-ABCDEABCDCDEAAAE-*Eco*RI. Similar analysis of the p3-6 sequence demonstrated that the monomeric configuration of the 15-mer is identical, with one of the triplicated group A monomers missing.

Relationship to other alpha satellite subsets. We have compared the sequences of alpha17 monomers to previously published alphoid monomer sequences from Old World monkeys (3, 19, 20) and humans (7, 10, 11, 18, 22, 33, 34). Monomers from chromosome 17 were from 57 to 82% identical in sequence to these other monomers. The significance of these numbers is not apparent, however, since no single monomer in p17H8 is representative of the higher-order repeat and since many published alphoid sequences do not reflect higher-order organization. Accordingly, we suggest that the most informative comparisons between alphoid subsets will come from comparison of complete sequences of the higher-order repeat units for any particular chromosomes. We have performed such an analysis for the p17H8 sequence and the alpha satellite subset (alphaX) previously characterized from the human X chromosome (28). The results of these comparisons are given in Table 3. For these comparisons, we have arbitrarily assigned the alpha17 *Eco*RI site as the start of the monomer array. For the alphaX pBamX7 sequence, the previously published monomers (aligned to start at the *Bam*HI site) were realigned to begin at a position 33 bp downstream from the *Bam*HI site, which corresponds to the position of the alpha17 *Eco*RI site.

The monomers of alphaX were previously shown to belong to five homology groupings (28). Based on the comparisons shown in Table 3, it is apparent that the homology groupings for both the alpha17 and alphaX higher-order repeat units coincide. For example, monomers 3 and 8 on the X belong to a homology group which, from the data in Table 3, is most closely related to monomers 5, 12, and 16 on chromosome 17 (group E). Thus, using the designations described above for the alpha17 subset, the monomer configuration of alphaX can be represented as -CDEABCDEABCD-. However, since alphaX and alpha17 higher-order repeat units are tandemly arranged on their respective chromosomes, the monomeric configuration can be viewed in any number of phases. In Fig. 4, alpha satellite from chromosomes X and 17 are each represented as tandem arrays, aligned for maximum agreement with respect to monomeric configuration. As highlighted in Fig. 4, the entire 12-mer configuration of alphaX is linearly conserved as part of the 16-mer alpha17 array. This shared 12-mer block comprises two adjacent, related pentamers (represented as EABCD) plus an additional two monomers. Although the chromosomal subsets of these two chromosomes have a conserved monomeric organization, the sequences in the common 12-mer blocks are only 85.5% identical, with al-

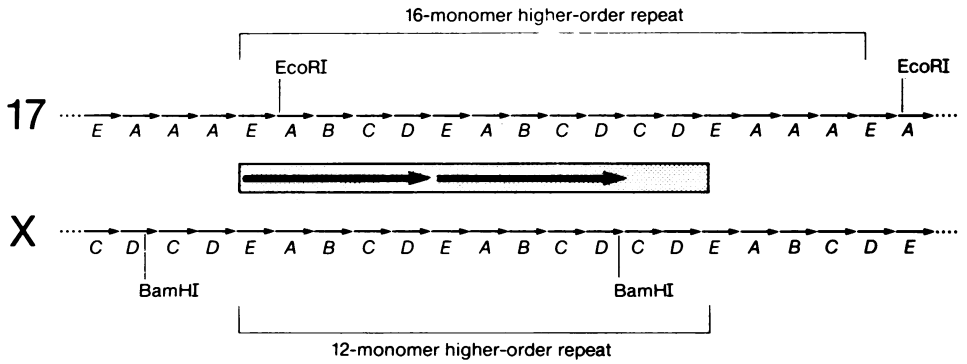

FIG. 4. Monomeric configurations of alpha17 and alphaX arrays. Tandem arrays of the higher-order repeat units of alpha17 (p17H8) and alphaX (pBamX7) are aligned for maximum agreement with respect to monomeric configuration. Small tandem arrows indicate alphoid monomers, and the designations A through E refer to the homology groupings described in the text. The *Eco*RI (alpha17) and *Bam*HI (alphaX) sites which define the higher-order repeat units are shown. A 12-mer block of conserved monomeric configuration is highlighted, as well as the pentamer blocks which constitute this region.

pha17-to-alphaX comparisons for individual monomers ranging from 79 to 93%. Thus, the alpha17 and alphaX subsets are related yet significantly diverged members of the human alpha satellite family.

To determine whether other chromosomal alpha satellite subsets are also closely related to the alpha17 subset, we performed Southern hybridization experiments with DNA from somatic cell hybrids containing individual human chromosomes (29). Under conditions of reduced hybridization stringency, the p3-6 probe hybridized strongly with alpha satellite of human chromosomes X, Y, 3, 4, and 21 (Fig. 5, left). Subsequent washing of the filter at high stringency, however, resulted in markedly reduced hybridization to alpha satellite from chromosomes Y, 3, 4, and 21 (Fig. 5, right). Hybridization with the human X chromosome, on the other hand, was at least partially resistant to this melting. Thus, the relationship between alpha17 and alphaX is apparent at both the nucleotide sequence and genomic DNA levels. Furthermore, these data indicate that alpha satellite subsets from at least four other human chromosomes are not as closely related.

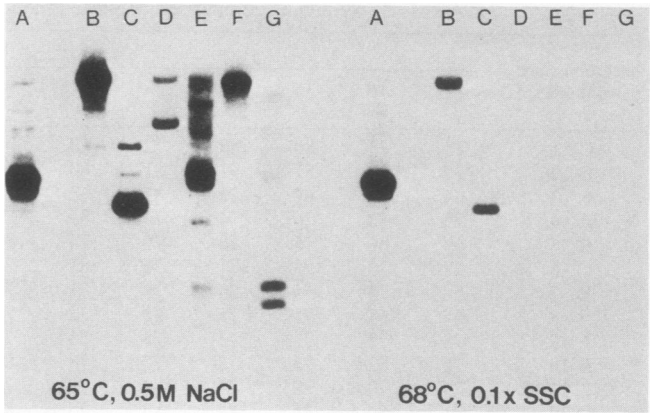

FIG. 5. Relationship between alpha17 and other chromosomal alpha satellite subsets. DNAs from somatic cell hybrids were probed with p3-6. Lanes: A, hybrid containing human chromosomes 5, 17, and 21 digested with *Eco*RI; B, human X-only hybrid, *Eco*RI; C, human X-only hybrid, *Bam*HI; D, human Y-only hybrid, *Hin*dIII; E, human 3-only hybrid, *Hin*dIII; F, human 4-only hybrid, *Eco*RI; G, human 21-only hybrid, *Eco*RI. The filter was first washed at reduced stringency (left panel) and subsequently rewashed at high stringency (right panel). Exposure times were equivalent.

DISCUSSION

In this report, we have described the isolation and characterization of representatives of alpha satellite DNA from human chromosome 17. Analysis of the structure and hybridization behavior of these repeated DNAs provides insights into the chromosome-specific organization and evolutionary history of this divergent human DNA family.

Definition of a higher-order repeat unit. Alpha satellite DNA is a tandemly repeated DNA family based primarily on a ~170- to 172-bp repeat length (4, 14, 16, 19, 20, 34). This length, the alphoid monomer, is the smallest repetitive unit of alpha satellite and exhibits no recognizable internal redundancy (34). Although alpha satellite comprises tandem monomers, this unit size does not necessarily correspond to the size of the presumed amplification domain during evolution. Like other tandemly repeated mammalian satellite DNAs, alpha satellite DNA can be characterized by longer-range periodicities marked by the presence of regularly spaced restriction enzyme recognition sites that define a higher-order repeat length (2, 4, 13, 14). Since definition of a higher-order repeat unit is dependent on the adventitious use or availability of a particular (and a priori unknown) restriction enzyme(s), conclusions regarding the equivalence of such operationally defined repeat lengths and repeat lengths of actual evolutionary significance are made with caution. Nonetheless, one prediction of the various models put forth to account for the amplification of higher-order tandem repeats is that independent copies of the most recent "true" amplification unit will be far closer in sequence than will be repeats within the amplified unit (2, 6, 23). Consistent with this notion, independent copies of the 12-mer repeat unit characteristic of alpha satellite from the human X chromosome are 98 to 99% identical in sequence, although the constituent monomers show 15 to 35% divergence among themselves (28). Similarly, independent cloned copies of a tetrameric higher-order repeat unit described by Gray et al. (7) are 98% identical, even though, again, the four constituent monomers show substantial (20 to 30%) variance.

Here, we have shown that the predominant higher-order repeat length of alpha satellite from human chromosome 17 is 2.7 kb in length, consisting of 16 tandem monomers. This 16-mer unit is itself tandemly arranged and is present in 500 to 1,000 copies per chromosome 17. Thus, based on a calculated length of chromosome 17 of 0.85×10^8 bp (25), the alpha17 family comprises, at a minimum, an estimated 1.5 to 3% of all chromosome 17 DNA. (We cannot rule out the possibility that there are other, related alpha satellite domains on this chromosome that have escaped detection with our probes under the conditions used.) Individual monomers within the 16-mer higher-order repeat unit are 65 to 94% identical in sequence to each other. In contrast, independent copies of the 2.7-kb repeat are >99% identical. This finding substantiates the view that the 2.7-kb 16-mer repeat represents a major unit of amplification of alpha satellite on chromosome 17.

Chromosome specificity. We and others have proposed that individual human chromosomes are each characterized by their own particular subset of alpha satellite (11, 15, 18, 29). Indeed, cloned alpha17 repeats (p3-6 and p17H8), when used as hybridization probes, hybridize predominantly with chromosome 17 DNA, failing to cross-hybridize appreciably with alpha satellite from other human chromosomes at high stringency (Fig. 2 and 5). This chromosome specificity is manifested not only by differential hybridization in filter assays, but also by different restriction enzyme periodicities and by different primary sequence. Thus, the current view of the human alpha satellite family is substantially more complex than indicated by early studies that emphasized the relatively simple 340-bp *Eco*RI alpha dimer (16, 34). Different alpha satellite subsets on different chromosomes can be recognized by different restriction enzymes. Furthermore, the size, sequence, and monomer suborganization of these higher-order repeat units can also vary considerably between chromosomes.

Relationship to other alpha satellite subsets. One consequence of the chromosome-specific organization of this DNA family is that detailed comparative analysis may reveal alpha satellite subsets that have a common ancestry. Indeed, in this study, we have shown that alpha satellite arrays on chromosomes 17 and X share a particular monomeric configuration based on a common pentameric core (Fig. 4). However, the conserved regions are only 85% identical in sequence. These subsets may therefore reflect independent amplifications of higher-order repeat units which have

evolved (diverged) from a common pentameric alphoid repeat.

The relationship between the alphaX and alpha17 subsets was confirmed by filter hybridization analysis of DNAs from certain individual human chromosomes (isolated in somatic cell hybrids) (Fig. 5). By assessing the thermostability of DNA-DNA hybrids, these experiments indicated that the chromosome 17 alpha satellite sequences were more closely related to sequences on the X chromosome than to sequences from chromosomes 3, 4, 21, or Y. Thus, alpha satellite on chromosomes 17 and X may define a subclass of human alpha satellite. In support of this notion, we have isolated alpha satellite repeats from human chromosomes 1 and 11 which also show striking similarities in monomeric configuration to the repeats on chromosomes 17 and X (unpublished data). On the other hand, this pentameric configuration shows no obvious relationship to either *Eco*RI alpha dimer or tetramer repeats (11, 22, 34) or the tetrameric repeat of Gray et al. (7). In addition, a 16-mer higher-order alpha satellite repeat unit from human chromosome 7 has been characterized and shows little evidence of a relationship to either the alphaX or alpha17 pentamer (unpublished data).

A simple interpretation of the present data would be that a common monomeric configuration, based on a pentamer, was initially present on several different human chromosomes, including 17 and X. Subsequently, these might have evolved independently to generate different higher-order repeat units on the different chromosomes, involving processes such as homologous, unequal crossing-over between misaligned repeats (see below). Finally, amplification and fixation of these higher-order repeat units could account for the contemporary chromosome-specific alpha satellite subsets.

Variability within the chromosome 17 alpha satellite subset. In this study, we have characterized two different higher-order repeats which belong to the alpha17 subfamily. Sequence determinations for both the 15-mer and the more prominent 16-mer repeat lengths revealed two levels of variation. First, the primary sequences were 2.2% diverged, a significantly higher level of divergence than observed among independent 16-mer clones. This degree of sequence divergence suggests that the 15-mer and 16-mer arrays reflect evolutionarily (as well as physically) distinct domains of alpha satellite on this chromosome. Second, the 15-mer repeat is precisely 1 monomer shorter than the 16-mer repeat (Fig. 1). This relative deletion involves one of three closely related monomers (monomers 13, 14, and 15) within the 16-mer. Because these three monomers are far closer in sequence to each other (91 to 94% identical) than to their flanking monomers 12 and 16 (67 to 73% identical; see Table 2), we propose that homologous, unequal crossing-over in this region could account for the formation of a 15-mer repeat from a 16-mer repeat (or vice versa). At present, we cannot distinguish which of the two repeat forms is the older. In some individuals, other repeat forms are apparent which correspond to higher-order repeat lengths, ranging from 14-mers to 17-mers (e.g., the 14-mer in Fig. 2A). Based on the evolutionary relationship described here for the 15-mer and 16-mer repeats, we would anticipate that these other repeat forms would also represent the products of relatively rare unequal crossing-over events during evolution and that these various repeat forms might differ from one another only by the number of group A monomers present between monomers 12 and 16 (of the 16-mer). Elsewhere, we have described additional variation in the alpha17 subset in the form of restriction fragment length polymorphisms detected as a 13-mer higher-order repeat unit (H. F. Willard, J. S. Waye, M. H. Skolnick, C. Schwartz, V. E. Powers, and S. B. England, Proc. Natl. Acad. Sci. USA, in press). Although the molecular basis for these polymorphisms has not yet been determined, this variation seems likely also to have arisen by unequal crossing-over.

It is worth noting that these proposed unequal exchanges all occur between tandem arrays misaligned within the higher-order repeat and thus involve reorganization of higher-order repeat structure. These events are therefore quite distinct from the presumed unequal crossing-over events which occur between higher-order repeats misaligned in register. These latter exchanges would be expected to result in changes in the copy number of the higher-order repeat but not its structure, and thus might be involved in fixation of the variant repeat lengths on a chromosome (23).

The "beginning" and "end" of monomers in tandem arrays. The beginning and end of individual monomers in a tandem array of monomers is arbitrarily set. For human alpha satellite monomers, there are 171 possible registers in which to read monomer sequences. For convenience, the start position of alphoid monomers has usually been set to coincide with the restriction enzyme site used to clone a particular fragment or to define a particular higher-order repeat unit. Thus, the African green monkey alphoid monomer sequence usually is presented from a *Hin*dIII site (19), while human *Eco*RI alpha dimers begin at the *Eco*RI site, out of register with the African green monkey repeat (16). Other primate alphoid sequences have been presented in yet other registers (3, 20).

For regularly repeating monomer configurations (e.g., -AAAAAAAA-, -ABABABABAB-, OR -ABCDEABCDE ABCDE-), the designation of arbitrary registers is without significance because of the precise tandem nature of the repeating monomers (one, two, or five monomers in the examples above). Any of the 171 registers would reveal the underlying monomeric configuration. Thus, unless a junction fragment from one end of the tandem array is available, which would allow empirical definition of the beginning of a monomer, any arbitrary register will suffice.

However, when obvious internal duplications or deletions which disrupt the regular pattern are evident, arbitrary registers are inappropriate and registers based on maximum homology between (for example) duplicated monomers may be more informative. For example, in the 16-mer higher-order repeat unit of alpha17, a region exists which appears to have three tandemly duplicated monomers (group A monomers 13, 14, and 15). To determine the register in which maximum homology is observed among these three monomers, the analysis shown in Fig. 6 was performed. Regions of the p17H8 sequence centered on monomers 13, 14, and 15 (indicated in Fig. 6A) were compared. Alignment of the nucleotide sequences of the three regions revealed relatively sharp boundaries separating areas showing different levels of homology (Fig. 6B). In the register selected to minimize mismatches among the three aligned sequences (indicated as register III in Fig. 6), 91% of the base positions are identical in the triplicated monomers, and all pairwise monomer comparisons are 94% identical. Register II represents a register based on the position of the *Eco*RI sites in the 16-mer (as in Fig. 3); it is clear that while this register significantly overlaps the register of greatest homology (thus allowing recognition of the apparent triplication [Table 2]), it also includes areas of substantial mismatch. Also indicated in Fig. 6 is a third register, in which monomers begin and end

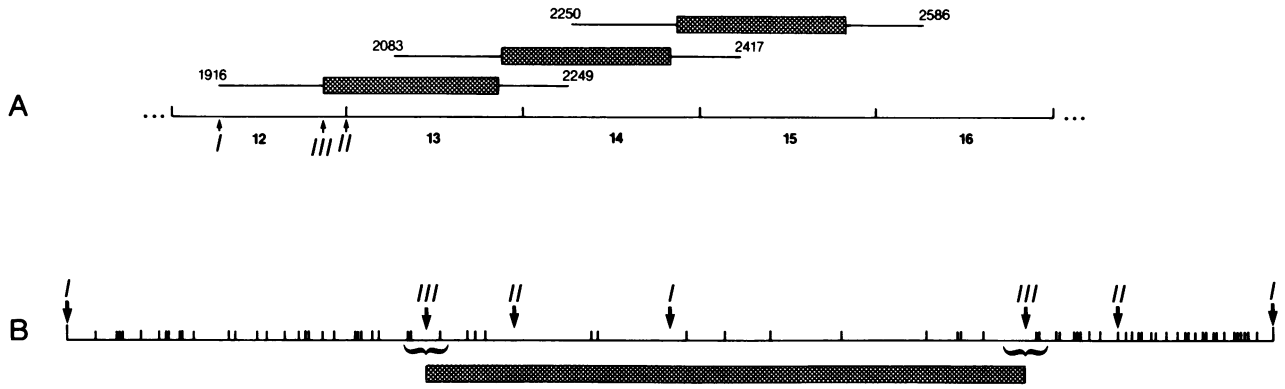

FIG. 6. Monomer register defining a triplication event. (A) Monomers 12 through 16 of p17H8 are represented with the apparent monomer triplication represented by monomers 13, 14, and 15. Sequence homologies were determined between the three regions bounded by positions 1916–2249, 2083–2417, and 2250–2586. (B) Homology profile for these regions. Vertical bars represent positions where there was at least one mismatch among the three sequences. Three registers are indicated: I, corresponding to the *Hin*dIII site of African green monkey; II, corresponding to the *Eco*RI site of p17H8; and III, a register determined by the least number of position mismatches. Placement of register III is approximate and can be shifted 5 to 10 bp in either direction. This register is highlighted by the 171-bp-long hatched regions in both A and B.

at the position corresponding to the African green monkey *Hin*dIII site (register I). When the sequences are viewed in this register, the triplication of group A monomers is not evident at all, since the aligned sequences include lengthy regions of mismatch involving monomers from different homology groups.

We suggest that regions showing greatest homology (defined by register III) represent those sequences involved in an unequal crossing-over event during evolution of the 15-mer or 16-mer repeat unit. By analogy with similar analyses of duplicated gene families, such as the globins and immunoglobulins (5), it seems reasonable to propose that the sharp transition between areas of markedly different levels of homology corresponds to the site of a presumed recombination breakpoint during unequal crossing-over between misaligned tandem arrays. It is perhaps of note that this site (at approximately positions 2010 through 2030 in Fig. 3) lies within a region noted previously to be significantly conserved among primate alpha satellite repeats (28). This site also corresponds to one of the three preferred binding sites in alpha satellite DNA for an AT-rich DNA-binding protein found in nuclear extracts of African green monkey cells (24, 26). While the significance of these associations is unclear, it will be of interest to examine other examples of homologous unequal crossing-over in alpha satellite arrays to determine whether such events are randomly distributed along the monomer or whether they preferentially occur at a limited number of sites.

ACKNOWLEDGMENTS

We thank D. Munroe and V. Powers for assistance with the hybrids and L. Creeper for contributions to the early stages of this project.

This work was supported by grants from the March of Dimes Birth Defects Foundation (1-995), the Hospital for Sick Children Foundation, and the Medical Research Council of Canada. H.F.W. is a research scholar of the Medical Research Council of Canada.

LITERATURE CITED

1. **Biggin, M. D., T. J. Gibson, and G. F. Hong.** 1983. Buffer gradients are 35-S label as an aid to rapid sequence determination. Proc. Natl. Acad. Sci. USA **80**:3963–3965.
2. **Brutlag, D. L.** 1980. Molecular arrangement and evolution of heterochromatic DNA. Annu. Rev. Genet. **14**:121–144.
3. **Donehower, L., C. Furlong, D. Gillespie, and D. M. Kurnit.** 1980. DNA sequence of baboon highly repeated DNA: evidence for evolution by non-random unequal crossovers. Proc. Natl. Acad. Sci. USA **77**:2129–2133.
4. **Donehower, L., and D. Gillespie.** 1979. Restriction site periodicities in highly repetitive DNA of primates. J. Mol. Biol. **134**:805–834.
5. **Ellison, J. W., and L. E. Hood.** 1983. Human antibody genes: evolutionary and molecular genetic perspectives. Adv. Hum. Genet. **13**:113–147.
6. **Gillespie, D.** 1977. Newly evolved repeated DNA sequences in primates. Science **196**:889–891.
7. **Gray, K. M., J. W. White, C. Costanzi, D. Gillespie, W. T. Schroeder, B. Calbretta, and G. F. Saunders.** 1985. Recent amplification of an alpha satellite DNA in humans. Nucleic Acids Res. **13**:521–535.
8. **Hagen, C. E., and G. J. Warren.** 1983. Lethality of palindromic DNA and its use in selection of recombinant plasmids. Gene **19**:147–151.
9. **Henikoff, S.** 1984. Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. Gene **28**:351–359.
10. **Jones, R. S., and S. S. Potter.** 1985. Characterization of cloned human alphoid satellite with an unusual monomeric construction: evidence for enrichment in HeLa small polydisperse circular DNA. Nucleic Acids Res. **13**:1027–1042.
11. **Jorgenson, A. L., C. J. Bostock, and A. L. Bak.** 1986. Chromosome-specific subfamilies within human alphoid repetitive DNA. J. Mol. Biol. **187**:185–196.
12. **Korneluk, R. G., F. Quan, and R. A. Gravel.** 1985. Rapid and reliable dideoxy sequencing of double-stranded DNA. Gene **40**:317–323.
13. **Kurnit, D. M.** 1979. Satellite DNA and heterochomatin variants: the case for unequal mitotic crossing over. Hum. Genet. **47**:169–186.
14. **Maio, J. J., F. L. Brown, and P. R. Musich.** 1977. Subunit structure of chromatin and the organization of highly repetitive DNA: recurrent periodicities and models for the evolutionary origins of repetitive DNA. J. Mol. Biol. **117**:637–656.
15. **Maio, J. J., F. L. Brown, and P. R. Musich.** 1981. Toward a molecular paleontology of primate genomes. I. The HindIII and EcoRI families of alphoid DNAs. Chromosoma **83**:103–125.
16. **Manuelidis, L., and J. C. Wu.** 1978. Homology between human and simian repeated DNA. Nature (London) **276**:92–94.
17. **Melton, D. A., P. A. Krieg, M. R. Rebagliati, T. Maniatis, K. Zinn, and M. R. Green.** 1984. Efficient in vitro synthesis of

biologically active RNA and RNA hybridization probes from plasmids containing a bacteriophage SP6 promoter. Nucleic Acids Res. **12**:7035–7056.

18. **Mitchell, A. R., J. R. Gosden, and D. A. Miller.** 1985. A cloned sequence p82H, of the alphoid repeated DNA family found at the centromeres of all human chromosomes. Chromosoma **92**:369–377.
19. **Rosenberg, H., M. F. Singer, and M. Rosenberg.** 1978. Highly repeated sequence of simiansimiansimiansimiansimiansimian. Science **200**:394–402.
20. **Rubin, C. M., P. L. Deininger, C. M. Houck, and C. W. Schmid.** 1980. A dimer satellite sequence in bonnet monkey DNA consists of distinct monomer subunits. J. Mol. Biol. **136**:151–167.
21. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
22. **Schmookler Reis, R. J., A. Srivastava, D. T. Beranek, and S. Goldstein.** 1985. Human alphoid family of tandemly repeated DNA: sequence of cloned tetrameric fragments and analysis of familial divergence. J. Mol. Biol. **186**:31–41.
23. **Smith, G. P.** 1976. Evolution of repeated DNA sequences by unequal crossing-over. Science **191**:528–535.
24. **Solomon, M. J., F. Strauss, and A. Varshavsky.** 1986. A mammalian high mobility group protein recognizes any stretch of six AT base pairs in duplex DNA. Proc. Natl. Acad. Sci. USA **83**:1276–1280.
25. **Southern, E. M.** 1982. Application of DNA analysis to mapping the human genome. Cytogenet. Cell Genet. **32**:52–57.
26. **Strauss, F., and A. Varshavsky.** 1984. A protein binds to a satellite DNA repeat at three specific sites that would be brought into mutual proximity by DNA folding in the nucleosome. Cell **37**:889–901.
27. **Vieira, J., and J. Messing.** 1982. The pUC plasmids, an M13mp7-derived system for insertion mutagenesis and sequencing with synthetic universal primers. Gene **19**:259–268.
28. **Waye, J. S., and H. F. Willard.** 1985. Chromosome-specific alpha satellite DNA: nucleotide sequence of the 2.0 kilobasepair repeat from the human X chromosome. Nucleic Acids Res. **13**:2731–2743.
29. **Willard, H. F.** 1985. Chromosome-specific organization of human alpha satellite DNA. Am. J. Hum. Genet. **37**:524–532.
30. **Willard, H. F., S. O. Meakin, L. C. Tsui, and M. L. Breitman.** 1985. Assignment of human gamma crystallin multigene family to chromosome 2. Somat. Cell Mol. Genet. **11**:511–516.
31. **Willard, H. F., and J. R. Riordan.** 1985. Assignment for the gene for myelin proteolipid protein to the X chromosome: implications for X-linked myelin disorders. Science **230**:940–942.
32. **Willard, H. F., K. D. Smith, and J. Sutherland.** 1983. Isolation and characterization of a major tandem repeat family from the human X chromosome. Nucleic Acids Res. **11**:2017–2033.
33. **Wolfe, J., S. M. Darling, R. P. Erickson, I. W. Craig, V. J. Buckle, P. W. Rigby, H. F. Willard, and P. N. Goodfellow.** 1985. Isolation and characterization of an alphoid centromeric repeat from the human Y chromosome. J. Mol. Biol. **182**:477–485.
34. **Wu, J. C., and L. Manuelidis.** 1980. Sequence definition and organization of a human repeated DNA. J. Mol. Biol. **142**:363–386.
35. **Yang, T. P., S. K. Hansen, K. K. Oishi, O. A. Ryder, and B. A. Hamkalo.** 1982. Characterization of a cloned repetitive DNA sequence concentrated on the human X chromosome. Proc. Natl. Acad. Sci. USA **79**:6593–6397.